

Roadmap

- Midterm total points
- Two corrections from last lecture
- r versus f
- F Statistics and migration
- Quantitative traits:
 - Partitioning variance
 - heritability and its hazards

Midterm total points

- ...only 95
- Will take this into account in interpreting the grades

Corrections

- I wrote: “If allele frequencies are not changing, why does homozygosity continue to go down over time?” which should be “heterozygosity”
- I wrote: “ $F_{ST} \approx \frac{t_S}{t_T}$ ”
- Someone correctly saw this doesn’t make any sense; it should be
- “ $F_{ST} \approx 1 - \frac{t_S}{t_T}$ ”

r versus f in a pedigree

- r is the expected proportion of alleles IBD in two relatives
- f is the chance that a hypothetical offspring of two relatives will have its two alleles IBD
- In straightforward cases, $f=r/2$
- Not straightforward:
 - Sex linked genes
 - Haplodiploids

Calculate r (in the simple case)

- Start at one individual and trace each possible path to the other
- Each parent-child link in a path is a factor of $1/2$, multiply to get probability along that path
- Add the paths together
- (Different paths are ones that go up to different relatives; for example full siblings have a relationship path through mother and a separate one through father)

Logic issues: r for haplodiploids

- What is r between father and daughter?
- From his point of view, 100% of his genes went to his daughter
- From her point of view, 50% of her genes came from her father
- What is r for:
 - Father-daughter?
 - Full sisters?
 - Full brother/sister?
 - Full brothers?

Why all these views of F_{ST} ?

- $\hat{F}_{ST} = \frac{\pi_B - \pi_W}{\pi_B}$ is how you would estimate it in practice
- The others are theoretical views of what it means:
 - Relationship of inbreeding coefficient within and between subpopulations
 - Relationship of mean coalescent depth within and between subpopulations
 - Relationship of variance within and between subpopulations

How is \hat{p} related to σ_T^2 ?

- $F_{ST} = \frac{\sigma_S^2}{\sigma_T^2} = \frac{\sigma_S^2}{\hat{p}(1-\hat{p})}$
- The denominator is a measure of how much variability is in the population as a whole
- The more variability (more even allele frequencies) in the overall population, the more differentiated the subpopulations can become

The dark side of F_{ST}

- $F_{ST} = \frac{\sigma_S^2}{\sigma_T^2} = \frac{\sigma_S^2}{\hat{p}(1-\hat{p})}$
- If \hat{p} is very large or small (one allele is very frequent), F_{ST} can't get big
- None the less, people attempt to give it an absolute interpretation
- Need to average across multiple loci to get meaningful results
- Still vulnerable to unexpected allele frequency spectrum (e.g. in growing/shrinking populations)

The dark side of F_{ST}

- $F_{ST} \approx \frac{\pi_B - \pi_W}{\pi_B}$
- Good luck if your sampling yields $\pi_B < \pi_W$, and it can!
- You can do better with ANOVA and its relatives, or with the coalescent methods to be described later
- ANOVA:
 - Decompose the total variation in the data into explanatory components (e.g. subpopulation structure)
 - Test against null hypothesis of no substructure
 - Program ARLEQUIN is a major tool for this

What is F_{ST} used for?

- In a model of stable population structure
 - Test for presence of structure
 - Estimate migration rate
- In a model of divergence from a common ancestor
 - Test for presence of structure
 - Estimate degree of divergence
- One further flaw: assumes one migration rate and one subpopulation size
- Better methods (ANOVA, AMOVA) are hard to code and explain!

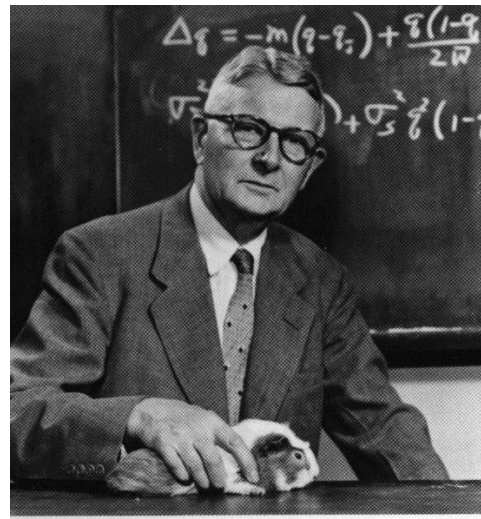
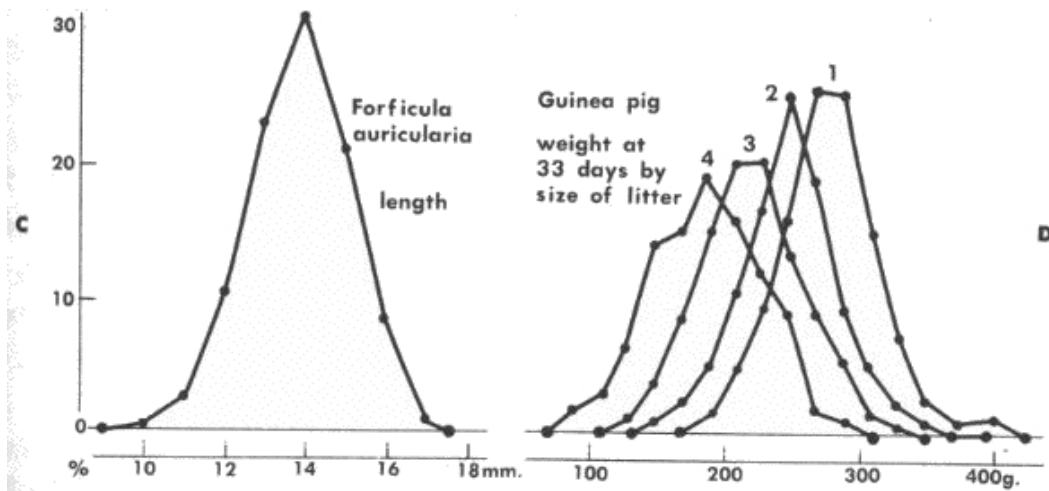
Next topic: Quantitative genetics

- Conflict between lab genetics and animal/plant breeders:
 - Lab genetics sees individual loci causing traits
 - Breeders see quantitative variation with no identifiable loci for many traits
- How can these views be reconciled?

Quantitative traits

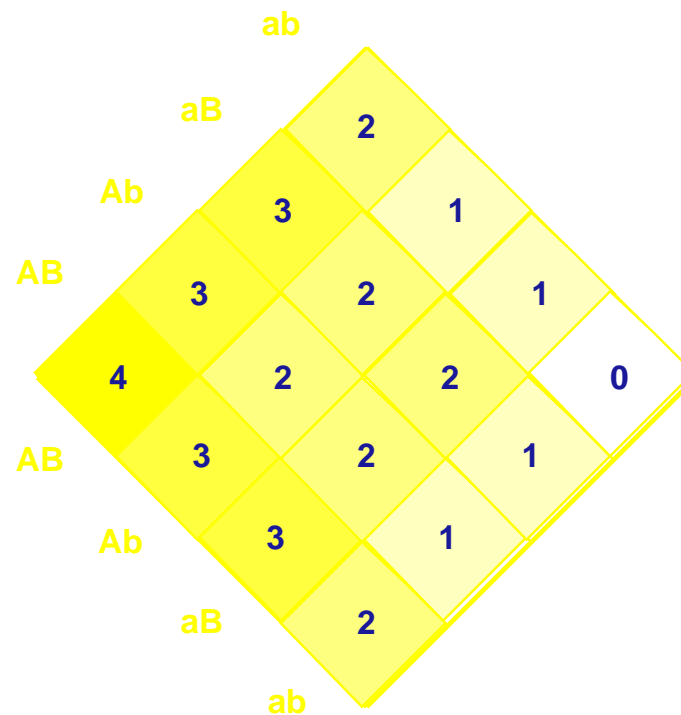
- Traits like height, weight, blood pressure, athletic performance, etc.
 - Likely to be polygenic with most alleles contributing only a small amount
 - Difficult to tackle on a gene by gene basis
 - Environment likely to be a major player
- Quantitative genetics abstracts the individual genes into a simplified model

Examples of quantitative traits



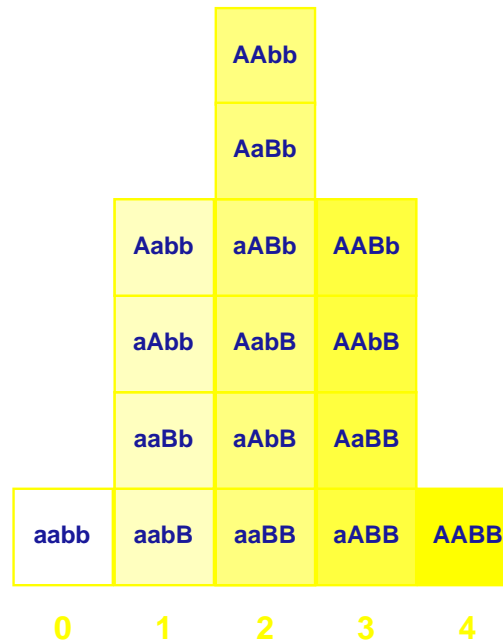
Multiple Mendelian characters → a quantitative trait

A random-mating population with two genes having
2 alleles each, at equal frequencies, symmetrically
affecting a quantitative character



Multiple Mendelian characters → a quantitative trait

The distribution of the genotypes and the quantitative character before artificial selection



Handling this type of bell-curve variability

- Estimate total variance in population for a trait:

$$V_T = \frac{\sum_i (x - \bar{x})^2}{i}$$

- x is a measurement of the trait on an individual
 - \bar{x} is the mean of the trait
 - i are the sampled individuals
- Quantitative methods try to partition this variance into genetic, environmental, and interaction terms
- (Should sound a bit familiar from F_{ST})

Why variance?

- If:
 - Multiple factors (f_1, f_2, \dots) affect a trait
 - Each is distributed as a normal (bell curve)
 - An individual's trait is the *sum* of these factors
- Then:
 - $V_T = V_{f1} + V_{f2} \dots$
- I wrote a small program to prove this to myself (available on request)

Partition the variance – a first attempt

- $V_T = V_G + V_E$
- V_G = genetic variance, V_E = environmental variance
- Pleasingly simple but not useful—why?
 - We'd like the “genetic” term to relate parents to offspring
 - Not all genetic variation can be used that way

Additive and non-additive genetic variation

- Assume variation in the trait purely due to one locus
- Additive variation:
 - AA = mean 10 kg, Aa = mean 8 kg, aa = mean 6 kg
 - Having an A allele increases your weight by 2 kg
 - It also increases your mean offspring weight by 2 kg
 - Creates a clear correlation between parent and offspring

Additive and non-additive genetic variation

- No additive variation:
- Assume A and a equally frequent
 - AA = mean 8 kg, Aa = mean 10 kg, aa = mean 8 kg
 - Having an A allele has no reliable effect on your weight
 - Passing one to your offspring has no reliable effect on offspring weight
 - No parent/offspring correlation
- Yet the trait is genetic! Not all genetic variation behaves the same....

A second try at decomposing variance

- $V_T = V_A + V_D + V_E + V_{GE}$
 - V_A – additive genetic variance
 - V_D – dominance genetic variance (all other genetic variance besides the additive component)
 - V_E – environmental variance
 - V_{GE} – covariance between genotype and environment

Genotype/environment interaction

- V_{GE} can arise when:
 - Your chance of having a genotype is correlated with your environment (ducks with lowland hemoglobin avoid mountains)
 - The effect of the environment depends on your genotype (highland hemoglobin only improves performance in mountains)
- “Common garden” experiment tries to remove these factors, BUT:
 - Results will not generalize back to wild population
 - Tendency to think a term we are “removing” is unimportant?
- Socioeconomic status is a common V_{GE} issue in humans

Appropriate scaling

- This theory is for genes with additive effects on phenotype
- Consider using $\log(\text{phenotype})$:
 - Loci of multiplicative effect create additive effect on $\log(\text{phenotype})$
 - Avoids unreasonable results like negative weight of an organism....
- Finding the “natural scale” of your trait would be even better, but is difficult

How to estimate the variance components?

- Pairs of relatives related through only one ancestor are easiest
- Phenotypic correlation between such relatives:
 - Depends on r
 - Depends on $h^2 = \frac{V_A}{V_T} = \frac{V_A}{V_A + V_D + V_E}$
- With only one shared ancestor, correlation = rh^2
 - Parent/child = $\frac{1}{2}h^2$
 - Half-sib = $\frac{1}{4}h^2$
- More complex formulas for multiple shared ancestors
- (Note that we are hoping V_{GE} will quietly go away—a particularly poor assumption for *relatives*)

Heritability

- h^2 is properly “narrow-sense heritability”
- (Broad-sense would be $\frac{V_G}{V_T}$ but is seldom used)
- “Of the phenotypic variation in this population, what proportion is due to additive genetic factors?”

What people actually want from heritability

- We'd like to answer questions like:
 - Is this trait genetic or environmental?
 - Can this trait possibly be affected by the environment?
 - What is the largest change in this trait possible via environmental manipulation? via breeding? via genetic manipulation?
- Heritability does not, and cannot, answer these questions
- Not clear they *have* answers

An illustrative paradox

- Consider traits essential for survival
- Heritability of such traits is generally *low*
- Why?

What is this actually good for?

- Predict one-generation phenotype response to selection on the trait
- $R = h^2 S$
- R is the response (the change in offspring relative to parents)
- S is the selection, measured as the difference between mean trait of breeding stock and mean trait of population

Friday

- Response to selection on a quantitative trait
- Long-term breeding experiments

One-minute responses

- Please:
 - Tear off a slip of paper
 - Give me one comment or question on something that worked, didn't work, needs elaboration, etc.